ELSEVIER
Journal of Cystic Fibrosis 7 (2008) 295–300
Journal of Cystic Fibrosis
www.elsevier.com/locate/jcf

ΔF508 mutation increases conformational flexibility of CFTR protein

G. Wieczorek [a,*], P. Zielenkiewicz [b]

[a] Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawinskiego 5a, 02-106 Warszawa, Poland
[b] Plant Molecular Biology Department, Warsaw University, Pawinskiego 5a, 02-106 Warszawa, Poland

Received 1 August 2007; received in revised form 21 November 2007; accepted 23 November 2007
Available online 29 January 2008

Abstract

Background: The deletion of Phe508 in the first nucleotide-binding domain of the CFTR protein is the most common mutation leading to cystic fibrosis.
Methods: We present a Molecular Dynamics study on the native and mutated domains, based on their recently published crystal structure.
Results: ΔF508 CFTR has much more conformational freedom compared to the wild-type, and exposes its hydrophobic interior to the solution.
Conclusions: The increased flexibility might be the reason for the recognition of mutated CFTR by the housekeeping proteins and its premature degradation. This, in turn results in reduction of population of functional channels at the epithelial cell surface and disease phenotype.

Keywords: Cystic fibrosis; CFTR; Molecular dynamics; Principal component analysis

1. Introduction

Cystic fibrosis is a lethal autosomal recessive genetic disorder correlated with abnormal Cl^- conductance at the plasma membrane of epithelial cells, resulting in the disruption of fluid and ion homeostasis. It is caused by defects in the cystic fibrosis transmembrane conductance regulator (CFTR), which is also a protein kinase A-regulated chloride ion channel [1–3]. The most common mutation responsible for cystic fibrosis is the deletion of residue Phe508 (ΔF508) in the first nucleotide-binding domain (NBD1) [4]. Normally, CFTR is assembled in the endoplasmatic reticulum, matures in the Golgi apparatus, and is delivered to the plasma membrane [5]. In the case of misfolded protein, it is degraded by the endoplasmic reticulum quality control [6] or by endosome-mediated internalization mechanisms [7]. The ΔF508 CFTR undergoes premature degradation much more often then wild-type protein. This results in a severe reduction in the population of functional channels at the epithelial cell surface. Also, because of regulatory function of CFTR, its lowered abundance leads to abnormal ionic transport of the cell. Experiments on CFTR done so far did not explain the great impact that the single point mutation has on the protein behaviour and further — on the cell's life. Structural and biophysical studies on complete human NBD1 domains failed to demonstrate significant changes of *in vitro* stability or folding kinetics in the presence or absence of the ΔF508 mutation [4]. Crystal structures show minimal changes in protein conformation (see Fig. 1). Changes in local surface topography at the site of the mutation, which is located in the region of NBD1 believed to interact with the first membrane spanning domain of CFTR, are considered to be the cause of altered interdomain interactions of the mutant protein [4]. However, there is another possible reason of abnormal CFTR trafficking and degradation that is to be investigated, rather complementary than contrary to the first one. Phenylalanine 508 interacts with surrounding aminoacids, contributing to the stiffness of entire region (see Fig. 2). Although the mutant protein retains its structure in the solution, it is possible, that the mutation slightly destabilizes the NBD1 domain, allowing it to cover much broader conformational space. As a result a hydrophobic part of the protein could be exposed, inducing detection of CFTR as misfolded and thus leading to

* Corresponding author. Tel.: +48 22 5925762.
E-mail address: gigo@ibb.waw.pl (G. Wieczorek).

1569-1993/$ - see front matter © 2007 European Cystic Fibrosis Society. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.jcf.2007.11.008

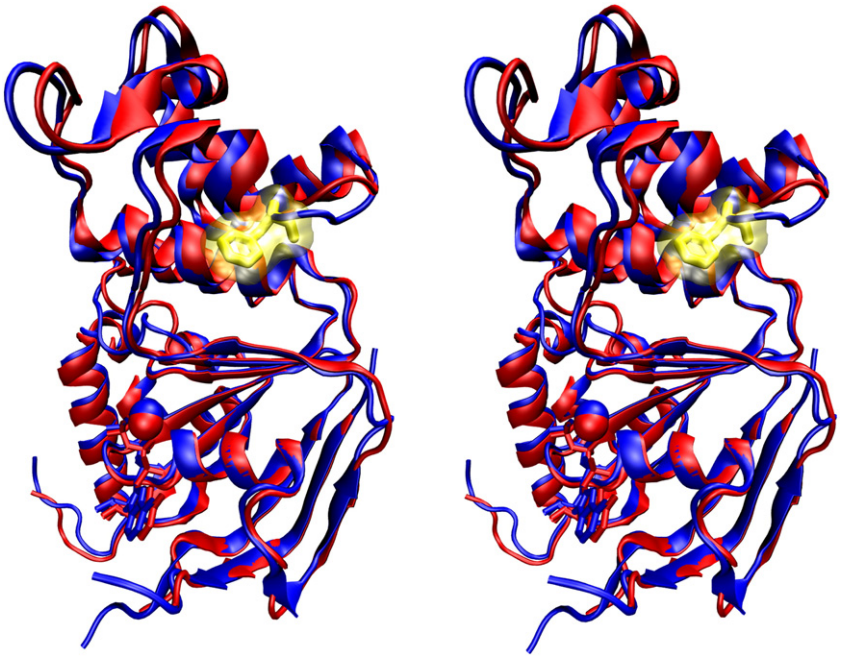

Fig. 1. Superposition of NBD1 domains of ΔF508 (1XMJ, red) and wild-type (2BBO, blue) CFTR in CrossEyes stereo cartoon representation. Phe508 selected as transparent surface and bonds. ATP molecules presented as bonds, Mg^{2+} ions as spheres (strongly overlapping here). All images prepared in VMD [24] and rendered in POV-Ray [31].

degradation of the protein. Checking whether such a hypothesis is justifiable requires comparative conformational analysis of CFTR with ΔF508 mutation and wild-type one. Since structures of the proteins were resolved crystalographically to a high resolution, they constitute a good input for Molecular Dynamics simulation studies, which were conducted to explore the flexibility of both proteins. In this paper it is shown that native and ΔF508 CFTR exhibit different behaviour. In particular, the mutant protein exposes more hydrophobic surface in a time frame allowing its detection by housekeeping proteins.

2. Methods

Crystalographically resolved structures of NBD1 domains from human ΔF508 CFTR mutant and CFTR without this deletion (it will further be called a wild-type protein for the sake of simplicity) — Protein Data Bank [8] entries 1XMJ and 2BBO [4], respectively — both complexed with ATP, were the starting points for the study. Both proteins contain the same 7 solubilizing mutations [4], so the only difference between examined parts of ΔF508 and "wild type" proteins was the existence of Phe508, which made the search for the influence of

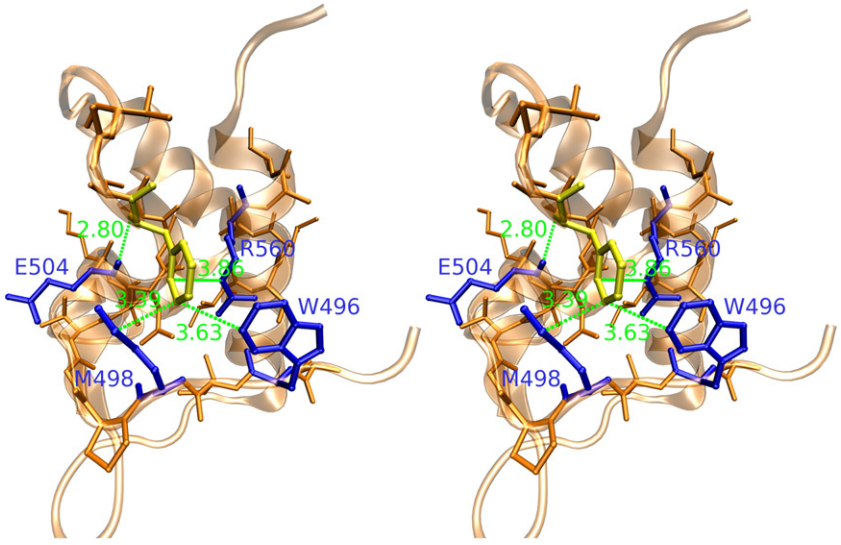

Fig. 2. Phenylalanine 508 (yellow) and its neighbourhood. Selected short interatomic distances of Phe508 heavy atoms with neighbouring aminoacids (blue) shown in green. Distances in Å.

this very mutation feasible. In both PDB entries there are regions for which the atomic coordinates have not been resolved by X-ray diffraction analysis. Coordinates of missing atoms in partially resolved residues were established by means of "profix" program from Jackal molecular modelling package [9] (1XMJ — 27 such atoms, 2BBO — 4 atoms). Next, unresolved loops missing in these structures were modelled with "loopy" program from above-mentioned package. This program generates several thousands of random backbone conformations of the missing loop, side chains are then assembled onto the backbone. Resulting loops are energy minimised. After potential energy evaluation, loops are further optimized and the best solution is chosen by means of the colony energy method [10]. In 1XMJ 4 loops built from 44 missing aminoacids in total were added, in 2BBO, respectively, 3 loops and 37 aminoacids. The correctness of the structures were checked by Procheck program [11], showing no major aberrations in the geometry of the proteins. The modelled parts of the proteins played only stabilising function. Proteins were converted into all-atom representations using OPLS-AA force field [12], which is known to correctly reproduce liquid state properties of macromolecules. Consecutive steps, except of the preparation of ATP topology, were done using Gromacs — Molecular Dynamics and analysis package [13]. Because at the time of described studies there were no topology parameters of ATP for OPLS-AA it was necessary to obtain such a data. PRODRG [14,15], a program for generating molecular topologies, have been used for producing ATP topology for GROMOS87 forcefield in Gromacs format. Information from this topology was manually combined with data contained in DNA/RNA records for OPLS force field in GROMACS format [16] based on existing OPLS and AMBER [17,18] parameters. The resulting topology was by no means checked or optimized, but during all simulations ATP remained bound to the protein and worked as a stabilizing ligand. The proteins were submerged in TIP4P water equilibrated for 20 ps at 300 K [19] in rhombic dodecahedral boxes. The walls of the boxes were not closer from the protein molecule than 20 Å. Ions were added to the solution to make the system electrically neutral. The system for wild-type protein consisted of 288 aminoacids (284 for ΔF508), 34,513 water molecules (31,961) out of which 36 were crystallization water molecules (135). Each system contained 10 sodium ions, one magnesium ion and one ATP particle. Periodic boundary conditions have been applied in all the simulations. For relaxing the solute–solvent contacts the systems underwent energy minimisation (steepest descent) with restrains on bonds length for 20,000 steps. The restrains were released and minimization was carried out for further 20,000 steps in order to eliminate residual strains. Next, 40 ps of Molecular Dynamics equilibration was performed. The integration of equations of motion in Molecular Dynamics simulations were done with leap-frog [20] algorithm, default in Gromacs. The solvent and the solute were (independently) weakly coupled to a temperature bath of 300 K with relaxation time 0.1 ps by means of Berendsen algorithm [21]. The Berendsen pressure coupling with coupling time equal to 0.8 ps and compressibility $4.5*10^{-5}$ bar^{-1} was applied. Long-range electrostatic interactions were computed using Particle Mesh Ewald (PME) electrostatics [22,23] with cubic interpolation. The relative strength of the Ewald-shifted direct potential at the cutoff was $1.0*10^{-5}$. Cutoff for van der Waals and Coulomb interactions were set to 9 Å. The maximum grid spacing for the fast Fourier transform (PME) was set to 1.2 Å. Finally, 20 ns of Molecular Dynamics were carried out for both proteins (timestep 1 fs). Every 1000 simulation steps a snapshot of the system was taken for further analysis. Trajectories obtained from Molecular Dynamics were analysed considering most striking differences in the conformational space covered by both proteins. The Principal Component Analysis (PCA) method, also called covariance analysis, has been applied for this

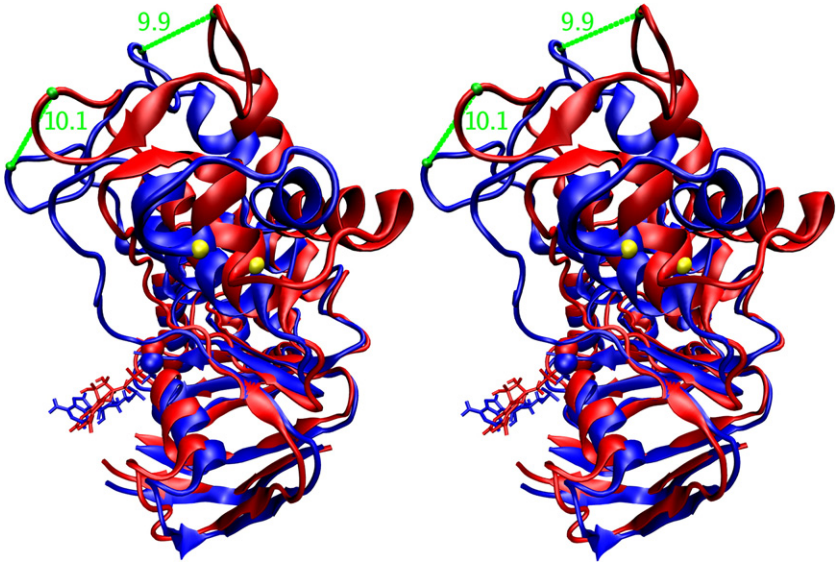

Fig. 3. Superposition of two frames from ΔF508 NBD1 trajectory. The frames shown gave extreme eigenvalues for the second important eigenvector, which signifies different behaviour of ΔF508 and wild-type protein. Yellow spheres indicate position of ΔF508 mutation. The distances between extreme locations of some backbone atoms located on loops in ABCα subdomain shown in green.

purpose [24]. This method can find correlated motions in molecules. Principal components that most significantly distinguish ΔF508 protein from the wild-type were further analysed with Visual Molecular Dynamics [25], WHATIF [26] and own programs. The PCA was also useful in assessing the efficiency of the conformational space sampling [27]. The subspace overlap of the first 8 eigenvectors of covariance matrices obtained for the halves of the trajectories of each simulation was calculated. For the above-mentioned calculations, backbone atoms of each protein, the same for wild-type and ΔF508, were selected for covariance matrices analysis. Namely, the common part of the backbone of crystalographically resolved regions of both proteins was taken (the modelled parts were neglected). Also, using the same method the whole trajectories of simulated proteins were compared. The trajectories were also subject of solvent accessible surface analysis. For every frame the surface area was calculated for every residue. Mean values and standard deviations were obtained by means of a fast online algorithm [28,29]. Both PCA and accessible surface analysis made it possible to determine hydrophobic patches exposed on the mutant protein, which might be responsible for recognition of ΔF508 CFTR as misfolded and leading to premature degradation.

3. Results

Calculation of the extreme projections along a trajectory for first few eigenvectors coming from covariance analysis simplified the task of finding the most bent or stretched conformations of the proteins (Fig. 3). The average solvent accessible surface of investigated part of proteins was 11,983 Å^2 for ΔF508 and 11,434 Å^2 for wild-type protein, while the average solvent accessible surface of hydrophobic residues were 2651 Å^2 and 2611 Å^2, and the values for frames of the highest degree of accessibility of hydrophobic residues are 3087 Å^2 and 3001 Å^2, respectively. The majority of the surface changes on ΔF508 protein takes its origins from the increase of conformational freedom gained by a linker between subdomains ABCα and ABCβ of NBD1 domain — aminoacids from 492 to 499, which disassociates from the body of the protein. The linker in the mutant protein remains for about 25% of the simulation time in the conformation not occurring during the whole simulation of wild-type protein. The surface of both subdomains of mutant protein becomes more variable then in wild-type case. Lack of stabilizing influence, which the Phe508 residue imposes on the linker, mainly on Met498 and Trp496

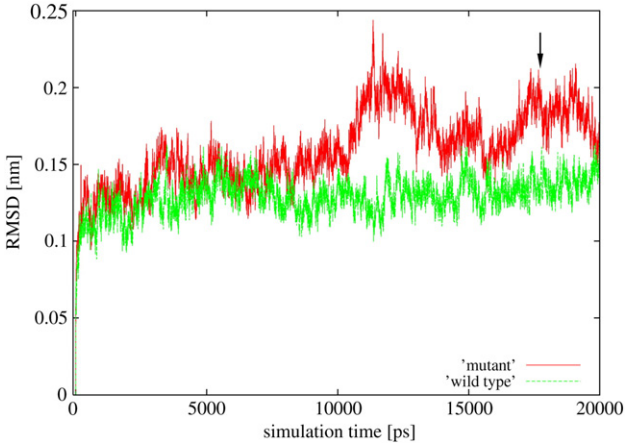

Fig. 5. Changes of the RMSD during the simulation. Whole trajectories were least-squares fitted to the starting structures using the backbone of the (most stable) beta-sheet ABCβ subdomain. The RMSD was then calculated for the whole backbone (not taking the modelled parts into account). Black arrow points to the simulation step visualized at Fig. 4 (frame 17738 of the simulation).

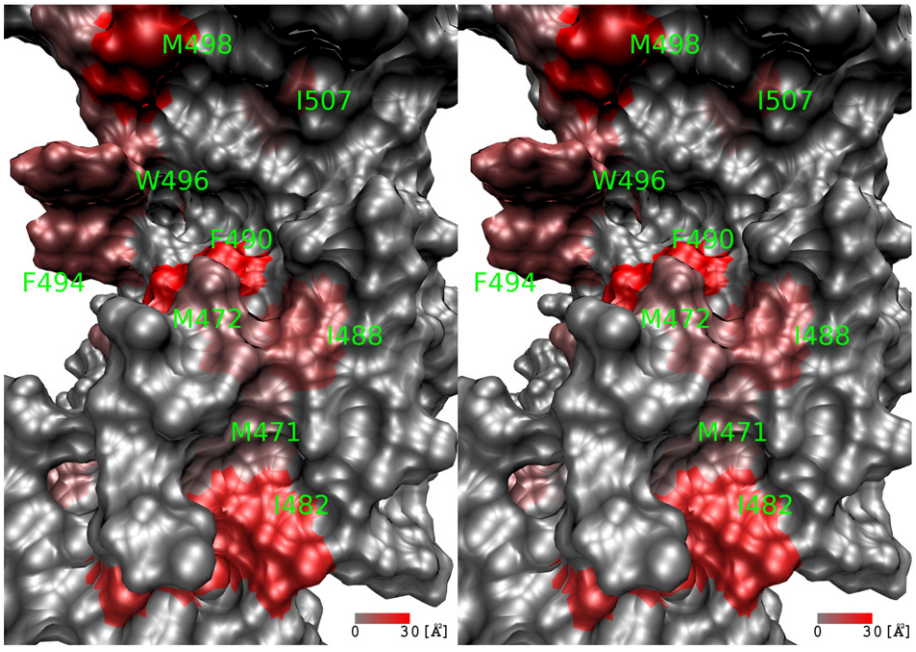

Fig. 4. An example of ΔF508 NBD1 hydrophobic surfaces exposed during simulation. The colour of aminoacid shows average solvent accessibility differences per residue between wild-type and ΔF508 NBD1 during whole simulation.

seems to play important role here. As the following effect, both the subdomains are allowed to bend (Fig. 3) more significantly than in the wild-type protein. The enhanced bending abilities and dissociation of the linker from the protein results in exposition of hydrophobic surfaces (Fig. 4). The increased flexibility of the ΔF508 protein is reflected also in higher deviations from the crystal structure conformation during the simulation (Fig. 5). As several cavities appear on the surface of the protein near the linker, it seems to be reasonable to hypothesise, that this place might be responsible for the recognition of the ΔF508 mutant by housekeeping proteins.

4. Discussion

The parts of the proteins that mostly differ in both simulated NBD1 types are residues 492–499 (linker between subdomains ABCα and ABCβ) and their surroundings. The above-mentioned linker of the mutant protein remained in the conformation different from what we can see in the wild-type for about 25% of the simulation time. There are, however, facts that make attempts of quantitative interpretation of the results futile. The normalized subspace overlap of covariance matrices built for the half of trajectory of the wild-type protein was 0.429, for ΔF508 protein — 0.501. Since overlap of value 1 indicates, that the sampled subspaces are identical, and value 0 — that they are orthogonal, the results suggest that it would be advantageous to carry out much longer molecular dynamics simulations or to use more efficient sampling method in order to more thoroughly capture the flexible nature of the proteins studied [27,30]. The subspace overlap of whole trajectories of wild-type and ΔF508 proteins was 0.46, which is a proof of significant similarity of the behaviour of both proteins. The necessity of more extensive sampling was also confirmed by RMSD measurements (Fig. 5). Thus, the results obtained should be treated as carrying qualitative, not quantitative, information as the latter would be significantly affected by statistical error. Also, the behaviour (time dependent conformation) of the proteins in cellular environment is certainly influenced by interactions with other macromolecules not taken into account in the presented simulations. Even non-specific crowding effects can play a major role here [32,33]. Despite of this, even in a relatively short simulation, it was possible to see at the first glance on the trajectories that NBD1 gains more conformational freedom due to the ΔF508 mutation. The ΔF508 mutant protein, although stable in the solution in terms of secondary structure, exhibits excessive motion between subdomains. This results in exposition of hydrophobic surfaces, by which the protein can be recognized as misfolded and thus undergoes premature proteolysis. Simulations performed in this work, although time-limited, allowed the exploration of the dynamic properties of two CFTR forms. Detection of conformational space and buried protein area changes helps in understanding the recognition of mutated CFTR as misfolded and suggests possible ways of interfering into this process. This approach is a basis for the development of small molecules, like peptidomimetics, that can either stabilise the protein not allowing for certain surface exposition, or preventing the housekeeping proteins from recognition of the already exposed surface. This in turn may lead to discoveries of new selective drugs [34].

Acknowledgements

This work was supported by the EU project LSHG-CT-2005-512044.

References

[1] Riordan JR, Rommens JM, Kerem B, et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 1989;245:1066–73.
[2] Rommens JM, Iannuzzi MC, Kerem B, et al. Identification of the cystic fibrosis gene: chromosome walking and jumping. Science 1989;245: 1059–65.
[3] Piserchio A, Fellows A, Madden DR, Mierke DF. Association of the cystic fibrosis transmembrane regulator with cal: structural features and molecular dynamics. Biochemistry 2005;44:16158–66.
[4] Lewis HA, et al. Impact of the deltaF508 mutation in first nucleotide-binding domain of human cystic fibrosis transmembrane conductance regulator on domain folding and structure. J Biol Chem 2005;280: 1346–53.
[5] Kopito RR. Biosynthesis and degradation of CFTR. Physiol Rev 1999;79:167–73.
[6] Ward CL, Omura S, Kopito RR. Degradation of CFTR by the ubiquitin-proteasome pathway. Cell 1995;83:121–7.
[7] Cheng J, Wang H, Guggino WB. Modulation of mature cystic fibrosis transmembrane regulator protein by the pdz domain protein cal. J Biol Chem 2004;279:1892–8.
[8] Berman HM, Westbrook J, Feng Z, et al. The protein data bank. Nucleic Acids Res 2000;28:235–42.
[9] Xiang Z, Honig B. Extending the accuracy limits of prediction for side-chain conformations. J Mol Biol 2001;311:421–30.
[10] Xiang Z, Cinque CS, Honig B. Evaluating conformational free energies: the colony energy and its application to the problem of loop prediction. PNAS 2002;99:7432–7.
[11] Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–91.
[12] Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. J Phys Chem B 2001;105:6474–87.
[13] Berendsen HJC, van der Spoel D, van Drunen R. Gromacs: a message-passing parallel molecular dynamics implementation. Comp Phys Comm 1995;91:43–56.
[14] van Aalten DM, Bywater R, Findlay JB, Hendlich M, Hooft RW, Vriend G. Prodrg, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. J Comput Aided Mol Des 1996;10:255–62.
[15] Schüttelkopf AW, van Aalten DM. Prodrg: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr D Biol Crystallogr 2004;60:1355–63.
[16] Golovin A, Polyakov N. DNA/RNA records for OPLS force field in GROMACS format; March 08, 2005. http://rnp-group.genebee.msu.su/3d/oplsa ff.html.
[17] Case DA, Cheatham TE, Darden T, et al. The amber biomolecular simulation programs. J Comput Chem 2005;26:1668–88.
[18] Ponder JW, Case DA. Force fields for protein simulations. Adv Protein Chem 2003;66:27–85.
[19] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–35.
[20] Hockney RW, Goel SP, Eastwood JW. Quiet high-resolution computer models of a plasma. J Comp Phys 1974;14:148–58.

[21] Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–90.
[22] Darden T, York D, Pedersen L. Particle mesh Ewald: an N∗log(N) method for Ewald sums in large systems. J Chem Phys 1993;98:10089–92.
[23] Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. J Chem Phys 1995;103:8577–93.
[24] Amadei A, Linssen AB, Berendsen HJ. Essential dynamics of proteins. Proteins 1993;17:412–25.
[25] Humphrey W, Dalke A, Schulten K. VMD — visual molecular dynamics. J Mol Graph 1996;14:33–8.
[26] Vriend G. WHAT IF: a molecular modeling and drug design program. J Mol Graph 1990;8:52–6.
[27] Hess B. Convergence of sampling in protein simulations. Phys Rev E 2002;65:031910.
[28] Welford BP. Note on a method for calculating corrected sums of squares and products. Technometrics 1962;4:419–20.
[29] Knuth DE. The art of computer programming. seminumerical algorithms, 3rd ed., 2. Addison-Wesley Longman Publishing Co., Inc.; 1997
[30] Hess B. Similarities between principal components of protein dynamics and random diffusion. Phys Rev E 2000;62:8438–48.
[31] Persistence of Vision Raytracer (POV-Ray), http://www.povray.org.
[32] Minton AP, Wilf J. Effect of macromolecular crowding upon the structure and function of an enzyme: glyceraldehyde-3-phosphate dehydrogenase. Biochemistry 1981;20:4821–6.
[33] Tokuriki N, Kinjo M, Negi S, et al. Protein folding by the effects of macromolecular crowding. Protein Sci 2004;13:125–33.
[34] Prongay AJ, et al. Discovery of the HCV NS3/4A protease inhibitor (1R,5S)-N-[3-amino-1-(cyclobutylmethyl)-2,3-dioxopropyl]-3-[2(S)-[[[(1,1-dimethylethyl)amino]carbonyl]amino]-3,3-dimethyl-1-oxobutyl]-6,6-dimethyl-3-azabicyclo[3.1.0]hexan-2(S)-carboxamide (Sch 503034) II. Key steps in structure-based optimization. J Med Chem 2007;50:2310–8.